



Универзитет у Београду
Математички факултет

СТАТИСТИЧКИ СОФТВЕР 1 (3В)

3. децембар 2025. године

1. У бази `Cars93` (пакет `MASS`) налазе се техничке и перформансне карактеристике 93 модела аутомобила.

(а) [2 поена] Формирати матрицу `StatMatr` у којој колоне одговарају променљивама `MPG.city`, `MPG.highway`, `Horsepower` и `Weight`, а врсте су редом:

- број опсервација без недостајућих вредности (`N`),
- медијана (`MEDIAN`),
- интерквartilни распон (`IQR`),
- распон (`R`), дефинисан као $R = \max - \min$.

(б) [1 поен] Нормализовати променљиву `Price` на интервал $[0, 1]$ применом формуле:

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)},$$

и колону именовати као `PriceNorm`.

(в) [2 поена] На основу `PriceNorm`, извршити класификацију аутомобила у три категорије:

- 1. `Low`: аутомобили код којих је $\text{PriceNorm} < 0.33$,
- 2. `Medium`: аутомобили код којих је $0.33 \leq \text{PriceNorm} \leq 0.66$,
- 3. `High`: аутомобили код којих је $\text{PriceNorm} > 0.66$.

Колону за класификацију аутомобила назвати `Class`, типа `factor`.

(г) [1 поен] Формирати листу са имена произвођача модела, где i -ти елемент листе садржи имена произвођача из i -тог кластера, $i = 1, 2, 3$.

(д) [2 поен] Нацртати дијаграм расипања (scatter plot) зависности `MPG.city` од `Weight`. На исти дијаграм додати праву линеарне регресије `MPG.city` у односу на `Weight`. Да ли постоји уочљива веза између потрошње у граду (`MPG.city`) и тежине аутомобила?

2. У бази `airquality` налазе се подаци о квалитету ваздуха у Њујорку 1973. године.

(а) [2 поен] Нацртати хистограм густине променљиве `Temp`. На исти графички приказ додати процењену функцију густине на основу података и криву густине нормалне расподеле са параметрима процењеним из података. Да ли емпиријска густина значајније одступа од нормалне?

(б) [1 поен] Имплементирати функцију `WindClass()` која на основу медијане променљиве `Wind` формира факторску колону `WindClass` са нивоима:

"Low" : (`Wind` < медијана), "High" : (`Wind` ≥ медијана).

Функција треба да врати листу са фреквенцијама класа и базу података проширену колоном `WindClass`.

(в) [2 поен] Имплементирати функцију `CleanAQ()` која из базе `airquality` уклања све редове у којима бар једна од нумеричких колона ("`Ozone`", "`Solar.R`", "`Wind`", "`Temp`") има недостајућу вредност. Функција враћа број уклоњених редова и очишћену базу.

(г) [2 поен] Редови у којима је бар једна вредност изван интервала

$$[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR]$$

сматрају се аутлајерима. Одредити број аутлајера у очишћеној бази.